# Take home exam
# Part II

***Convex Unconstrained and Constrained Optimization***

March 28, 2022

Gloria del Valle Cano
gloria.valle@estudiante.uam.es

---

**Problem 1.** (1 point) We have worked out the elementary version of Lagrange multipliers assuming that from $g(x, y) = 0$ we can find a function $y = h(x)$ such that $g(x, h(x)) = 0$.
But sometimes what we get is that there is an $h$ such that $g(h(y), y) = 0$. Rewrite the Lagrange multiplier analysis in the lecture slides under this assumption.

For $f, g : \mathbb{R}^2 \to \mathbb{R}$ consider the following minimization problem

$$\min f(x, y) \text{ s.t. } g(x, y) = 0.$$

Assuming the **Implicit Function Theorem** holds, we can find a function $x = h(y)$ s.t. $g(h(y), y) = 0$ and, thus, we can write

$$f(x, y) = f(h(y), y) = \Psi(y).$$

At a minimum $y^*$ with $x^* = h(y^*)$ we thus have

$$0 = \Psi'(y^*) = \frac{\partial g}{\partial x}(x^*, y^*)h'(y^*) + \frac{\partial g}{\partial y}(x^*, y^*). \tag{1}$$

But since $g(h(y), y) = 0$, we also have

$$0 = \frac{\partial g}{\partial x}(x^*, y^*)h'(y^*) + \frac{\partial g}{\partial y}(x^*, y^*) \implies h'(y^*) = -\frac{\frac{\partial g}{\partial y}(x^*, y^*)}{\frac{\partial g}{\partial x}(x^*, y^*)}. \tag{2}$$

Putting together (1) and (2) we arrive at

$$0 = \frac{\partial f}{\partial y}(x^*, y^*)\frac{\partial g}{\partial x}(x^*, y^*) - \frac{\partial f}{\partial x}(x^*, y^*)\frac{\partial g}{\partial y}(x^*, y^*).$$

That is, at $(x^*, y^*)$, $\nabla f \perp \left(-\frac{\partial g}{\partial y}, \frac{\partial g}{\partial x}\right)$ and, since $\left(-\frac{\partial g}{\partial y}, \frac{\partial g}{\partial x}\right) \perp \nabla g$, we have $\nabla f \| \nabla g$ i.e. $\nabla f(x^*, y^*) = -\lambda^* \nabla g(x^*, y^*)$ for some $\lambda^* \neq 0$.

Thus, for the **Lagrangian**

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda g(x, y),$$

we have a minimum $(x^*, y^*)$ there is a $\lambda^* \neq 0$ s.t.

$$\nabla \mathcal{L}(x^*, y^*, \lambda^*) = \nabla f(x^*, y^*) + \lambda^* \nabla g(x^*, y^*) = 0.$$

**Problem 2.** (3 points) We want to solve the following constrained minimization problem:

$$\begin{aligned} \min \quad & f(x, y) = x^2 + 2xy + 2y^2 - 3x + y \\ \text{s.t.} \quad & x + y = 1, \\ & x \geq 0, y \geq 0. \end{aligned}$$

Argue first that $f$ is convex and then:

- Write its Lagrangian with $\alpha, \beta$ the multipliers of the inequality constraints.
- Write the KKT conditions.
- Use them to solve the problem. For this consider separately the $(\alpha = \beta = 0), (\alpha > 0, \beta = 0), (\alpha = 0, \beta > 0), (\alpha > 0, \beta > 0)$ cases.

First of all, we verify that $f$ is convex because its Hessian matrix is positive semidefinite, or equivalently its eigenvalues are non-negative. The Hessian matrix is

$$\mathbf{H}_f = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} \\ \dfrac{\partial^2 f}{\partial y \partial x} & \dfrac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix},$$

whose eigenvalues are $\lambda_1 = 3 + \sqrt{5}$ and $\lambda_2 = 3 - \sqrt{5}$, both positive. As a result, we determine $f$ is convex.

Now we write its Lagrangian, with $\alpha$ and $\beta$ as multipliers of the inequality constraints and $\lambda$ as equality constraint multiplier.

$$\mathcal{L}(x, y, \lambda, \alpha, \beta) = x^2 + 2xy + 2y^2 - 3x + y + \lambda(x + y - 1) - \alpha x - \beta y.$$

Assuming that the hypothesis of the KKT conditions theorem hold, the resulting KKT conditions on a local minimum $(x^*, y^*)$ are the following:

$$\begin{aligned} 0 &= \frac{\partial \mathcal{L}}{\partial x}(x^*, y^*, \lambda, \alpha, \beta) \implies \lambda + 2x^* + 2y^* - 3 - \alpha = 0, \\ 0 &= \frac{\partial \mathcal{L}}{\partial y}(x^*, y^*, \lambda, \alpha, \beta) \implies \lambda + 2x^* + 4y^* + 1 - \beta = 0, \\ 0 &= \alpha x^*, \\ 0 &= \beta y^*. \end{aligned}$$

Then, we use them to solve the problem, considering the four possible cases below:

- Case $\alpha = \beta = 0$.

$$0 = \frac{\partial \mathcal{L}}{\partial x} = \lambda + 2x + 2y - 3$$
$$0 = \frac{\partial \mathcal{L}}{\partial y} = \lambda + 2x + 4y + 1$$

  From both expressions we get $2x+2y-3 = 2x+4y+1 \implies 2y = -4 \implies y = -2 \not\geq 0$, so this point does not satisfy this constraint. Then, $x = 1 - y = 3$ and $\lambda = -1$. So we have that $(\mathbf{3}, -\mathbf{2})'$ **is not a feasible solution**.

- Case $\alpha > 0, \beta = 0$. When $\alpha > 0$, $x = 0$, so $y = 1$ and $\lambda = -5$. Finally, $\alpha = -6 \not> 0$. Therefore we have that $(\mathbf{0}, \mathbf{1})'$ **is not a feasible point**.

- Case $\alpha = 0, \beta > 0$. When $\beta > 0$, $y = 0$, so $x = 1$ and $\lambda = 1$. Finally, $\beta = 4 > 0$. Therefore we have a **feasible KKT point** on $(\mathbf{1}, \mathbf{0})'$.

- Case $\alpha > 0, \beta > 0$. This implies $x = y = 0$, so we have a contradiction because $x + y \neq 1$. This means that $(0, 0)'$ **is not a feasible point**.

Given all points found, we can determine that our optimal solution is $(\mathbf{1}, \mathbf{0})'$, with an optimal value of $\mathbf{f}((\mathbf{1}, \mathbf{0})) = -\mathbf{2}$.

**Problem 3.** (1 point) Let $f : S \subset \mathbb{R}^d \to \mathbb{R}$ be a convex function on the convex set $S$ and we extend it to an $\tilde{f} : \mathbb{R}^d \to \mathbb{R}$ as:

$$\begin{aligned} \tilde{f}(x) &= f(x) \text{ if } x \in S. \\ &= +\infty \text{ if } x \notin S. \end{aligned}$$

Show that $\tilde{f}$ is a convex function on $\mathbb{R}^d$. Assume that $a + \infty = \infty$ and that $a \cdot \infty = \infty$ for $a > 0$.

We say that $S$ is a **convex set** if for all $x, x' \in S$ and $\lambda \in [0, 1]$,

$$\lambda x + (1 - \lambda) x' \in S.$$

Let $x, x' \in S$ and $\lambda \in [0, 1]$, so here we cover two cases:

- *First case.* If $x, x' \in S$:

$$\tilde{f}(\lambda x + (1-\lambda)x') = f(\lambda x + (1-\lambda)x') \leq \lambda f(x) + (1-\lambda) f(x') = \lambda \tilde{f}(x) + (1-\lambda)\tilde{f}(x').$$

  Where the first equality holds given that $S$ is convex. The inequality holds because $f$ is convex. And, the last equality raises from the definition of $\tilde{f}$.

- *Second case.* If $x \notin S$ or $x' \notin S$, we have that

$$\tilde{f}(\lambda x + (1-\lambda)x') \leq \lambda \tilde{f}(x) + (1-\lambda)\tilde{f}(x') = +\infty,$$

  because $\tilde{f}(y) \leq +\infty$, $\forall y \in \mathbb{R}^d$.

Since both cases satisfy convexity definition holds, we conclude this function is convex.

**Problem 4.** (2 points) Prove **Jensen's inequality**: if $f$ is convex on $\mathbb{R}^d$ and $\sum_1^k \lambda_i = 1$, with $0 \leq \lambda_i \leq 1$ we have for any $x_1, \dots, x_k \in \mathbb{R}^d$

$$f\left(\sum_1^k \lambda_i x_i\right) \leq \sum_1^k \lambda_i f(x_i)$$

*Hint: just write $\sum_1^k \lambda_i x_i = \lambda_1 x_1 + (1-\lambda_1)$ for an appropriate $v$ and apply repeatedly the definition of a convex function. Start with $k = 3$ and carry on.*

We proceed using an inductive procedure:

- If $k = 1$ then $\lambda = 1$, so we simply have $f(x_1) \leq f(x_1)$, which is true, and nothing to prove. If $k = 2$ we have the definition of the convexity of $f$:

$$\lambda_1 + \lambda_2 = 1,\ \lambda_1, \lambda_2 \geq 0 \implies f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2).$$

- Considering $k = 3$, given $\lambda_1 + \lambda_2 + \lambda_3 = 1$, and $\lambda_2, \lambda_3 > 0$, we have the following inequality

$$\begin{aligned}
f\left(\lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3\right) &= f\left(\lambda_1 x_1 + (1-\lambda_1)\left(\frac{\lambda_2}{\lambda_2+\lambda_3} x_2 + \frac{\lambda_3}{\lambda_2+\lambda_3} x_3\right)\right) \\
&\leq f(\lambda_1 x_1) + f\left((1-\lambda_1)\left(\frac{\lambda_2}{\lambda_2+\lambda_3} x_2 + \left(1 - \frac{\lambda_2}{\lambda_2+\lambda_3}\right) x_3\right)\right) \\
&\leq \lambda_1 f(x_1) + (1-\lambda_1)\left(\frac{\lambda_2}{\lambda_2+\lambda_3} f(x_2) + \left(1 - \frac{\lambda_2}{\lambda_2+\lambda_3}\right) f(x_3)\right) \\
&= \sum_{i=1}^{3} \lambda_i f(x_i),
\end{aligned}$$

  where we applied the case $k = 2$ twice.

- We assume the statement its true for $k$ and consider $k+1$ points $x_1, \dots, x_{k+1}$, with coefficients $\lambda_1, \dots, \lambda_{k+1} \geq 0$, $\sum_{i=1}^{k+1} \lambda_i = 1$. The evaluation of the linear combination can be decomposed as

$$f\left(\sum_{i=1}^{k+1} \lambda_i x_i\right) = f\left((1-\lambda_1)\left(\sum_{i=2}^{k+1} \frac{\lambda_i}{1-\lambda_1} x_i\right) + \lambda_1 x_1\right).$$

  Using this, it is straightforward to use the Jensen's inequality on $x_1$ and $\sum_{i=2}^{k+1} \frac{\lambda_i}{1-\lambda_1} x_i$ with coefficients $\lambda_1$ and $1-\lambda_1$ respectively. That is,

$$f\left(\sum_{i=1}^{k+1} \lambda_i x_i\right) \leq (1-\lambda_1) f\left(\sum_{i=2}^{k+1} \frac{\lambda_i}{1-\lambda_1} x_i\right) + \lambda_1 f(x_1).$$

We may notice that $\sum_{i=2}^{k+1} \frac{\lambda_i}{1-\lambda_1} x_i$ verifies the inductive hypothesis, thus,

$$f\left(\sum_{i=2}^{k+1} \frac{\lambda_i}{1-\lambda_1} x_i\right) \leq \sum_{i=2}^{k+1} \frac{\lambda_i}{1-\lambda_1} f(x_i).$$

Finally,

$$f\left(\sum_{i=1}^{k+1} \lambda_i x_i\right) \leq (1-\lambda_1) \sum_{i=2}^{k+1} \frac{\lambda_i}{1-\lambda_1} f(x_i) + \lambda_1 f(x_1) = \sum_{i=1}^{k+1} \lambda_i f(x_i).$$

**Problem 5.** (3 points) Prove that the following function is convex

$$\begin{aligned} f(x) &= x^2 - 1, \qquad && |x| > 1 \\ &= 0 && |x| \leq 1 \end{aligned}$$

and compute its proximal. Which are the fixed points of this proximal?

We note that $f$ can be seen as the maximum of two functions $f(x) = \max\{0, x^2 - 1\}$. Both of these functions are convex. Then, we are going to show that the maximum of two convex functions is also convex.

Let $m$ and $n$ be two convex functions and $h(x) = \max\{m(x), n(x)\}$. Given $\lambda \in [0,1]$ and $x, y \in \mathbb{R}$ we aim to show that

$$h(\lambda x + (1-\lambda)y) \leq \lambda h(x) + (1-\lambda) h(y).$$

On one hand, it is clear that

$$m\left(\lambda x + (1-\lambda)y\right) \leq \lambda m(x) + (1-\lambda) m(y) \leq \lambda h(x) + (1-\lambda) h(y),$$

where in the first inequality comes from the fact that $m$ is convex and the second from the definition of $h$. The same inequalities holds for $n$. Given that both functions are upper bounded by the same value, the maximum is also upper bounded by this value, so $h$ is convex. Using this auxiliary result, $f$ is convex.

Now we compute the proximal operator of $f$ as

$$\operatorname{prox}_f(x) = \arg\min_z f(z) + \frac{1}{2}(z-x)^2 = \arg\min_z h(z)$$

with

$$h(z) = \begin{cases} z^2 - 1 + \frac{1}{2}(z-x)^2 & |z| > 1 \\ \frac{1}{2}(z-x)^2 & |z| \leq 1 \end{cases} \tag{3}$$

If the minimizer is attained at $|z| \leq 1$, then clearly $z = x$, meaning that $\operatorname{prox}_f(x) = x$ for $|x| \leq 1$. If it is attained at $|z| > 1$, we have

$$0 = h'(z) = 3z - x \implies z = \frac{1}{3}x,$$

which implies that $\operatorname{prox}_f(x) = \frac{1}{3}x$ for $|x| > 3$. The remaining values of the proximal must be studied separately: the only possible minimizers are the points of non differentiability of (3). That is $-1$ and $1$ with

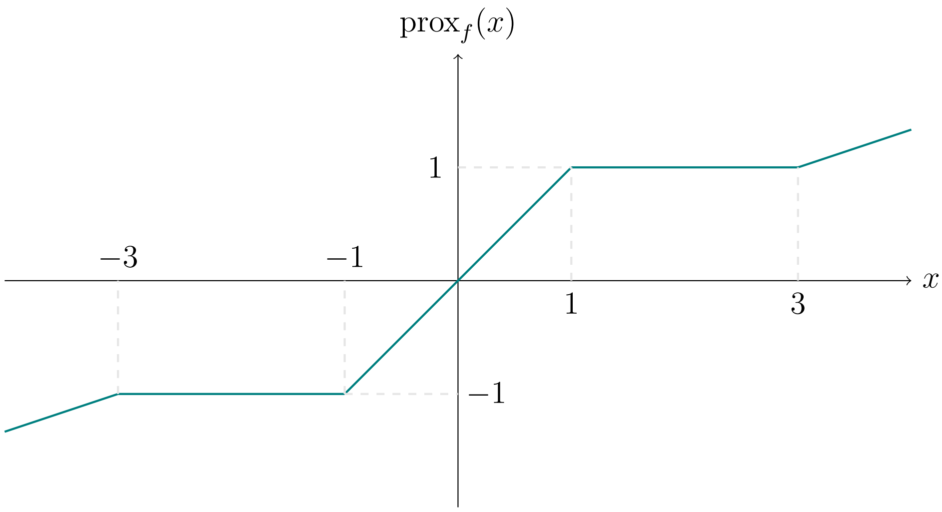
$$h(-1) = \frac{1}{2}(-1-x)^2 \quad \text{and} \quad h(1) = \frac{1}{2}(1-x)^2.$$

- When $-3 \leq x < -1$, the function is minimized at $z = -1$, that is, the proximal is $\operatorname{prox}_f(x) = -1$.
- When $1 < x \leq 3$, the function is minimized at $z = 1$, that is, the proximal is $\operatorname{prox}_f(x) = 1$.

As a result, the proximal is

$$\operatorname{prox}_f(x) = \begin{cases} \frac{x}{3} & x \in (-\infty, -3), \\ -1 & x \in [-3, -1), \\ x & x \in [-1, 1], \\ 1 & x \in (1, 3], \\ \frac{x}{3} & x \in (3, \infty), \end{cases}$$

Furthermore, we illustrate $\operatorname{prox}_f(x)$ for a better comprehension.




Finally, we observe that our fixed points of this proximal are those that verify $|x| \leq 1$.